

Segmentation of MS lesions using entropy-based EM algorithm and Markov random fields

Ahmad Bijar¹, Mahdi Mohamad Khanloo¹, Antonio Peñalver Benavent², Rasoul Khayati¹

¹Department of Biomedical Engineering, Shahed University, Tehran, Iran;

²Department de Estadística, Matemáticas e Informàtica, Universidad Miguel Hernández, Elche, Spain.

Email: ahmad_bijar@live.com; mahdi.khanloo@gmail.com; a.penalver@umh.es; khayati@shahed.ac.ir

Received 25 June 2011; revised 28 July 2011; accepted 8 August 2011.

ABSTRACT

This paper presents an approach for fully automatic segmentation of MS lesions in fluid attenuated inversion recovery (FLAIR) Magnetic Resonance (MR) images. The proposed method estimates a gaussian mixture model with three kernels as cerebrospinal fluid (CSF), normal tissue and Multiple Sclerosis lesions. To estimate this model, an automatic Entropy based EM algorithm is used to find the best estimated Model. Then, Markov random field (MRF) model and EM algorithm are utilized to obtain and upgrade the class conditional probability density function and the *a priori* probability of each class. After estimation of Model parameters and *a priori* probability, brain tissues are classified using bayesian classification. To evaluate the result of the proposed method, similarity criteria of different slices related to 20 MS patients are calculated and compared with other methods which include manual segmentation. Also, volume of segmented lesions are computed and compared with gold standard using correlation coefficient. The proposed method has better performance in comparison with previous works which are reported here.

Keywords: Gaussian Mixture Model; EM; Entropy; Markov Random Field; Multiple Sclerosis

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) permits the non-invasive detailed visualization of internal anatomical structures. The segmentation of MR images into anatomical tissues, fluids, and structures is an area of interest in MRI, and is increasingly being used to assess the progression of the disease and to evaluate the effect of drug therapy, supplementing traditional neurological disability scales such as the extended disability status scale (EDSS) [1]. Multiple Sclerosis (MS) is a chronic,

inflammatory, demyelinating disease of the Central Nervous System (CNS) that primarily affects the white matter (WM) of the central nervous system. MS lesions exhibit hypersignals in T2 weighted and hyposignals in T1 weighted, with respect to normal white matter intensities. Previous research has shown that the FLAIR (Fluid Attenuated Inversion Recovery) sequence contains the most distinctive lesion-healthy tissue differentiation for segmentation of white matter lesions. The radiological criteria for MS include the number of lesion on the MRI, their locations and their sizes, and these quantitative information is also crucial for studying the progression of MS lesions and the effect of drug treatments. To make a quantitative analysis of the brain scans of the patient, human experts are required to identify the multiple sclerosis lesions that are present in those scans as manual segmentation which is extremely time consuming. So, automatic and semi-automatic methods for segmentation of MS lesions are recommended and have been investigated extensively. There are many proposed approaches, automatic and semi-automatic, for segmentation of brain into different tissues, including MS lesions. These approaches include a variety of methods such as statistical, fuzzy, neural networks, and fuzzy neural networks. In this paper, we develop a fully automatic method for segmentation of MS lesions that requires no training, atlas, or thresholding steps. Our method can be divided into two main sections: at first, intensities of brain pixels are modeled using a gaussian mixture model which consists of three kernels as CSF, normal tissue and MS lesions classes. This step starts with only one kernel and uses an entropy based EM algorithm to estimate three kernels as three mentioned classes in an automatic manner which does not need initial values for parameter estimation. After estimation of the model, we would be able to classify brain pixels by knowing *a priori* probabilities of the classes. The next step is obtaining these *a priori* probabilities with no training and atlas. So, a MRF model and EM algorithm

are applied to update and attain *a priori* probabilities and means and variances of each class.

2. METHODS

2.1. Patients and MR Imaging

The proposed method is evaluated on a dataset which is obtained and used in [2]. This dataset contains 16 female and 4 male with average age of 29 ± 8 years old, this dataset is selected according to the revised Mc Donald criteria 2005 [3]. All images were acquired according to full field MRI criteria of MS [3] in T2-weighted (T2-w), T1-weighted (T1-w), Gadolinium enhanced T1-weighted, and FLAIR in axial, sagittal and coronal surfaces. We selected the FLAIR images, especially axial ones, with lesions in deep, periventricular, subcortical, and cortical white matters (supratentorial lesions). More lesion load and higher accuracy of FLAIR in revealing of these MS lesions were the reason for this selection [4]. Each image volume (patient data) consisted of averagely 40 slices with a 256×256 scan matrix. The pixel size was 1mm^2 , and the slice thickness was 3mm without any gap.

2.2. Manual Segmentation of MS Lesions and Brain Segmentation

The segmentation of MS lesions was performed manually by neurologist and radiologist in Flair images with visual inspection of corresponding T1-w and T2-w images. These manually segmented images were used as Gold standard [5] to evaluate the performance of proposed method. To evaluate the proposed method, different types of images which have different lesion volumes were applied. Also, the brain segmentation was performed using a fully automatic object-oriented approach [6]. This method was based on the regional-spatial characteristics of brain in MR images. At first, original image is converted to a binary image. Then, morphological opening on the binary image is performed and tiny regions are eliminated. Three rectangular masks showing the cerebral regions are produced and the regions in the binary image which have overlap with these rectangles are preserved and, the rest are eliminated. Final mask is generated by dilation of selected regions and filling tiny holes. Finally, an image, which includes only cerebral tissues, is obtained by applying the resulted mask on the original image.

3. PROBLEM FORMULATION

3.1. Gaussian Mixture Model

A finite mixture model $p(x)$ is the weighted sum of $M > 1$ components in \mathbb{R}^n for $n \geq 1$

$$p(x) = \sum_{m=1}^M \pi_m p(x|m) \quad \forall x \in \mathbb{R}^n \quad (1)$$

where $\pi_m \in (0,1)$ ($m=1,2,\dots,M$) corresponds to the

weight of each component which satisfies $\sum_{m=1}^M \pi_m = 1$. For the Gaussian mixtures model, each component density $p(x|m)$ is a Gaussian probability density given by

$$p(x|\theta_m) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma_m)^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(x-\mu_m)^T \Sigma_m^{-1} (x-\mu_m)\right\} \quad (2)$$

where T denotes the transpose operation, μ_m is the mean vector and Σ_m is the covariance matrix which is assumed positive definite. Here we encapsulate these parameters into a parameter vector, writing the parameters of each component as $\theta_m = (\mu_m, \Sigma_m)$, to get $\Theta = (\pi_1, \pi_2, \dots, \pi_m, \theta_1, \theta_2, \dots, \theta_m)$. **Eq.1** can be rewritten as

$$p(x|\Theta) = \sum_{m=1}^M \pi_m p(x|\theta_m) \quad (3)$$

If we knew the component from which x came, then it would be simple to determine the parameters Θ . Similarly, if we knew the parameters Θ , we could determine the component that would be most likely to have produced x . The difficulty is that we know neither.

3.2. Bayesian Classification

Bayesian Classification is a probabilistic technique of pattern recognition and is based on the principle of Bayes decision theory [7], given in **Eq.4** below

$$P(\omega_j|x) = \frac{p(x|\omega_j) \cdot P(\omega_j)}{p(x)} \quad (4)$$

where, x is a given feature vector, ω_j denotes a class, or state of nature, $P(\omega_j)$ is the prior probability of class ω_j , $p(x)$ is prior probability of the feature vector x , $p(x|\omega_j)$ is *a posteriori* probability, which a feature vector should be classified as belonging to class ω_j , $P(\omega_j|x)$ is the conditional probability that a feature vector occurs in a given class ω_j . For the approach here, the feature x shall consist of one component, intensity of brain pixels. The quantity $p(x)$ is known as the evidence, and serves only as a scale factor, such that the quantity in **Eq.4** is indeed a true probability, with values between zero and one. So, the maximum a posteriori (MAP) estimate of **Eq.4** is used as below

$$P(\omega_j|x) = p(x|\omega_j) \cdot P(\omega_j) \quad (5)$$

According to Bayesian theory [8], the feature vector x is classified to ω_j of which the *a posteriori* probability given x is the largest between the classes.

$$p(x|\omega_j) \cdot P(\omega_j) = \max_i \{p(x|\omega_i) \cdot P(\omega_i)\} \quad x \in \omega_j \quad (6)$$

Bayes decision rule is optimal in the sense of minimization of the probability of error. It is quite obvious that such an Ideal Bayesian solution can be used only if distributions $p(x|\omega_j)$, and the *a priori* probabilities $P(\omega_j)$ are known. In the context of classification of brain tissue, the probability models are not known, and therefore, must be approximated. The performance of the Bayesian classifier is directly related to how well these distributions can be modeled.

3.3. EM Algorithm

The Expectation-Maximization (EM) algorithm [9] is a general-purpose iterative algorithm for maximum likelihood (ML) estimation. The EM algorithm consists of two steps, E-step (expectation) and M-step (maximization) which are alternately used until the $\log \prod_{i=1}^n p(x_i|\Theta)$ converges to a local optimum. Let $X = \{x_i\}, i = 1, \dots, n$ be a set of n realizations of a random d-dimensional vector χ with a pdf $p(x_i)$. Thus the parameterized pdf can be written as a combination of pdfs of the M components $w_m (m = 1, \dots, M)$ characterizing the finite mixture model

$$p(x_i|\Theta) = \sum_{m=1}^M \pi_m p(x_i|\theta_m) \quad (7)$$

The likelihood can be expressed as

$$L(\Theta) = \prod_{i=1}^n p(x_i|\Theta) = \prod_{i=1}^n \sum_{m=1}^M \pi_m p(x_i|\theta_m) \quad (8)$$

E-step: the data χ are assumed to be incomplete and the complete data set $Y = (\chi, Z)$ is determined by estimating the set of variables $Z = \{Z_1, Z_2, \dots, Z_M\}$, where each Z_M is an N-dimensional vector $[z_m^1, z_m^2, \dots, z_m^N]^T$. The log likelihood of the complete data Y is

$$\log p(Y|\Theta) = \sum_{i=1}^N \sum_{m=1}^M z_m^i \log [\pi_m p(x^i|\theta_m)] \quad (9)$$

where

$$z_m^i = P(m|x^i, \Theta^t) = \frac{\pi_m^t p(x^i|\theta_m^t)}{\sum_{l=1}^M \pi_l^t p(x^i|\theta_l^t)} \quad (10)$$

is the posterior probability and Θ^t is the parameter estimate obtained after t iterations.

M-step: in this step, the parameters Θ^{t+1} are determined according to the estimate of the variables z_m^i . For Gaussian mixture models this corresponds to reestimating the π_m^{t+1} , the μ_m^{t+1} and Σ_m^{t+1} for each m according to:

$$\pi_m^{t+1} = \frac{1}{N} \sum_{i=1}^N z_m^i \quad (11)$$

$$\mu_m^{t+1} = \frac{\sum_{i=1}^N z_m^i x_i}{\sum_{i=1}^N z_m^i} \quad (12)$$

$$\sum_m^{t+1} = \frac{\sum_{i=1}^{t+1} z_m^i (x_i - \mu_m^{t+1})(x_i - \mu_m^{t+1})^T}{\sum_{i=1}^N z_m^i} \quad (13)$$

4. ENTROPY

In the information theory, information entropy is a measurement of the system uncertain degree. The term of entropy in the theory of information was first introduced by Shannon, in 1948, in [10].

4.1. Shannon Entropy

Let $X = \{x_1, \dots, x_n\}$ be a discrete random variable with values in S and probability mass function $p(x)$, the information (or uncertainty) of each possible event is

$$I(x_i) = -\log p(x_i) \quad (14)$$

The mean of $I(X)$ is introduced as Shannon entropy and is denoted by $H(X)$

$$H(X) = E(I(X)) = -\sum_{x \in S} p(x) \log p(x) \quad (15)$$

$H(X)$ varies from zero to $\log(|S|)$ zero meaning that there is no uncertainty, whereas $\log(|S|)$ is reached when all elements of X have equal probabilities, in this case, the uncertainty is at its maximum [11].

4.2. Rényi Entropy

Rényi further expanded the concept of Sannon entropy, and defined the $q (q \geq 0, q \neq 1)$ order Rényi [12] entropy of probability density function $p(x)$ as

$$H_q(X) = \frac{1}{1-q} \log \sum_{x \in S} p(x)^q, \quad q \neq 1 \quad (16)$$

Rényi entropy is a non-increasing function of q and It can be verified, by applying the l'Hôpital rule that Shannons entropy is the Rényi entropy of order 1.

5. ENTROPY ESTIMATION

In this paper, Rényi entropy [12] is selected as entropy measure and estimation of this type of entropy which will be express, is based on Leonenko and *et al.* [13].

Let $X \in \mathbb{R}^n$ be a random vector with probability measure μ having the density p . The proposed method by Leonenko estimates H_q from a sample of N independent and identically distributed (i.i.d.) random variables $X_1, X_2, \dots, X_N, N \geq 2$ based on the nearest neighbor distances in the sample. This work is resumption of the method which estimates H_1 and proposed by Kozachenko and Leonenko [14].

Leonenko's method estimates I_q , **Eq.17**, for $q \neq 1$ through the computation of conditional moments of nearest-neighbor distances.

$$I_q = I_q(p) = \mathbb{E}\{p^{q-1}(x)\} = \sum_{x \in S} p(x)^q, \quad q \neq 1 \quad (17)$$

5.1. Estimators

Suppose that $X_1, X_2, \dots, X_N, N \geq 2$, are i.i.d. with a probability measure μ having a density p with respect to the Lebesgue measure. Let $\rho(x, y)$ denote the Euclidean distance between two points x, y of \mathbb{R}^m . For a given sample X_1, \dots, X_N and a given X_i in the sample, from the $N-1$ distances $\rho(X_i, X_j), j=1, \dots, N, j \neq i$, the order statistics $\rho_{1, N-1}^{(i)} \leq \rho_{2, N-1}^{(i)} \leq \dots \leq \rho_{N-1, N-1}^{(i)}$ are formed, so that $\rho_{k, N-1}^{(i)}$ is the k th nearest-neighbor distance from X_i to some other X_j in the sample, $j \neq i$. I_q in Eq.18 is estimated for $q \neq 1$, by

$$\hat{I}_{N,k,q} = \frac{1}{N} (\zeta_{N,i,k})^{1-q} \tag{18}$$

where

$$\zeta_{N,i,k} = (N-1) C_k V_m (\rho_{k, N-1}^{(i)})^m$$

$$V_m = \pi^{m/2} / \Gamma(m/2 + 1)$$

(V_m is the volume of the unit ball $\beta(0,1)$ in \mathbb{R}^m)

$$C_k = [\Gamma(k) / \Gamma(k+1-q)]^{1/(1-q)}$$

So, H_q in Eq.17 is estimated as below

$$\widehat{H}_{N,K,q} = \log(\hat{I}_{N,k,q}) / (1-q) \tag{19}$$

5.2. Maximum Entropy

Suppose that we have a group of variables which have obtained from different types of distributions and have equal variance. According to the second Gibbs theorem [15], among all of them, Gaussian variables have maximum entropy. Therefore, by using this theorem, we would be able to understand that if the modeling of the variable with a gaussian kernel is acceptable or not. This procedure is done through the comparing of the theoretical maximum entropy and the real one of the underlying data. Let X be a d dimensional variable, the theoretical maximum entropy is computed as below

$$H_{\max}(X) = \frac{1}{2} \log[(2\pi)^d |\Sigma|] \tag{20}$$

where, Σ is covariance of observations. There would be no difference between the theoretical maximum entropy and the real one. If the data were truly Gaussian. We have used this property to determine the worst estimated kernel in the mixture.

6. SELECTION AND SPLIT OF THE WORST ESTIMATED KERNEL

For every kernel in the mixture, the theoretical maximum entropy and the real one of the underlying data is computed and evaluated in the normalized weighted

form [16] to find the worst estimated kernel. The worst estimated kernel would be selected through

$$k^* = \arg \max_k \left\{ \pi_k \frac{H_{\max}(k) - H_{\text{real}}(k)}{H_{\max}(k)} \right\} \tag{21}$$

where, k^* indicates the worst estimated kernel, $H_{\text{real}}(k)$ is the real entropy of the data under the k th kernel, and $H_{\max}(k)$ is the maximum entropy of the k th kernel. Suppose that a special kernel has been selected to split, during the split step, overall dispersion and two new covariance matrices have to be constant and positive define, respectively. If the k th component had to split into the k_1 th and k_2 th components, the first two moments, the mean vectors and the covariance matrices should satisfy the following split equations which are proposed by Richardson and Green [17]

$$\begin{aligned} \pi_k &= \pi_{k_1} + \pi_{k_2} \\ \pi_k \mu_k &= \pi_{k_1} \mu_{k_1} + \pi_{k_2} \mu_{k_2} \\ \pi_k (\Sigma_k + \mu_k \mu_k^T) &= \pi_{k_1} (\Sigma_{k_1} + \mu_{k_1} \mu_{k_1}^T) + \pi_{k_2} (\Sigma_{k_2} + \mu_{k_2} \mu_{k_2}^T) \end{aligned} \tag{22}$$

Solving the split equations is an ill-posed problem because the number of the equations is less than the number of the unknowns. To solve this ill-posed problem, we have used the method which is proposed by Dellaportas and Papageorgiou in [18]. They have used the spectral decomposition of the current covariance matrix, and the original problem is replaced by estimating the new eigenvalues and eigenvectors of the new covariance matrices without the previously cited constraint. Let $\Sigma_k = V_k \Lambda_k V_k^T$ the spectral decomposition of the covariance matrix Σ_k , with V_k , the eigenvector matrix, $\Lambda_k = \text{diag}(\lambda_k^1, \dots, \lambda_k^d)$ a diagonal matrix containing the eigenvalues of Σ_k with increasing order, and d the problem dimension. Let also D be a $d \times d$ rotation matrix with orthonormal unit vectors as columns. D is built by generating its lower triangular matrix independently from $d(d-1)/2$ different uniform $U(0,1)$ densities. The proposed split operation is given by

$$\begin{aligned} \pi_{k_1} &= \mu_{k_1} \pi_k \\ \pi_{k_2} &= (1 - \mu_{k_1}) \pi_k \\ \mu_{k_1} &= \mu_k - \left(\sum_{i=1}^d u_2^i \sqrt{\lambda_k^i} V_k^i \right) \sqrt{\frac{\pi_{k_2}}{\pi_{k_1}}} \\ \mu_{k_2} &= \mu_k + \left(\sum_{i=1}^d u_2^i \sqrt{\lambda_k^i} V_k^i \right) \sqrt{\frac{\pi_{k_1}}{\pi_{k_2}}} \\ \Lambda_{k_1} &= \text{diag}(u_3) \text{diag}(t - u_2) \text{diag}(t + u_2) \Lambda_k \\ \Lambda_{k_2} &= \text{diag}(t - u_3) \text{diag}(t - u_2) \text{diag}(1 + u_2) \Lambda_k \end{aligned}$$

$$\begin{aligned} \mathbf{V}_{k_1} &= D\mathbf{V}_k \\ \mathbf{V}_{k_2} &= D^T\mathbf{V}_k \end{aligned} \quad (23)$$

where $\mathbf{1}$ is a $d \times 1$ vector of ones, and

$u_1, u_2 = (u_1^1, u_2^2, \dots, u_2^d)^T$, and $u_3 = (u_3^1, u_3^2, \dots, u_3^d)^T$ are $2d + 1$ random variables needed to build priors, means, and eigenvalues for the new component in the mixture. They are calculated as

$$\begin{aligned} u_1 &\sim \beta(2, 2) & u_2^1 &\sim \beta(1, 2d) \\ u_2^j &\sim U(-1, 1) & u_3^1 &\sim \beta(1, d) & u_3^j &\sim U(0, 1) \end{aligned} \quad (24)$$

with $j = 2, \dots, d$, and $U(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ denoting beta and uniform distributions, respectively. The expressions for μ_1 and μ_2 establish that the new means are obtained by moving across the axes formed by the eigenvectors with sizes $u_1^1 \sqrt{\lambda_k^1}$ and $u_2^2 \sqrt{\lambda_k^2}$. We must keep u_1^1 positive but let u_2^2 vary in $[-1, 1]$ the interval because the moves in both directions across the large eigenvector are achieved via expressions for μ_1 and μ_2 , but the moves across the shorter axis require both positive and negative u_2^2 .

7. ESTIMATION OF PROBABILITY DENSITY FUNCTION

This section describes the proposed method for fitting Gaussian mixture models in depth. The algorithm starts with only one kernel, whose initial parameters are given by the whole samples. As is shown in [19], if we denote $y_1, y_2, \dots, y_N \in \mathbb{R}^d$ a set of samples of a pdf, the ML hypothesis for the mean and covariance matrix are

$$\begin{aligned} \mu_1 &= \frac{1}{N} \sum_{i=1}^N y_i \\ \Sigma_1 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_1)^T \end{aligned} \quad (25)$$

Furthermore, the initial value for the first prior π_1 is 1. EM algorithm computes a fitted model which consists only one kernel by these initial values. Whereas the model has only one kernel, this kernel should be split and replaced by two other ones. This work is done by Eq.23. The new model with two kernels has to fit on samples. EM algorithm with initial values which are obtained by previous split, is applied for this reason. Now, the fitted model has two kernels as two classes and it is less than our interested number of kernels which is three, csf, normal tissue and MS lesions. So, the worst fitted kernel among these two kernels has to be selected and split. As mentioned before, the worst fitted kernel has big difference between the theoretical maximum entropy and the real one against other kernels. So, we could select the worst kernel by means of maximum entropy and real one using Eq.21.

The worst fitted kernel is replaced by two new kernels through split (Eq.23) and the other kernel is remained with no variation. To find the final fitted model which has three kernels, another EM algorithm is applied, the initial parameters for EM algorithm are obtained by split (split of the worst fitted kernel) and from previous EM algorithm (for remained kernel). Figure 1 shows the Entropy-based EM algorithm to fit the best fitted mixture to brain samples.

8. MARKOV RANDOM FIELDS

An advantage of MRF models [20] is the use of neighborhood information to improve the *a priori* probabilities $P(\omega)$ [21]. The intuition behind the MRF model is that most pixels belong to the same class as their neighbors, and it is a powerful tool to describe the class assigning or labeling dependence between adjacent pixels.

Suppose a digital image bases on a $M \times N$ lattice, so image describes as

$$S = \{s = (x, y) | 1 \leq x \leq M, 1 \leq y \leq N\} \quad (26)$$

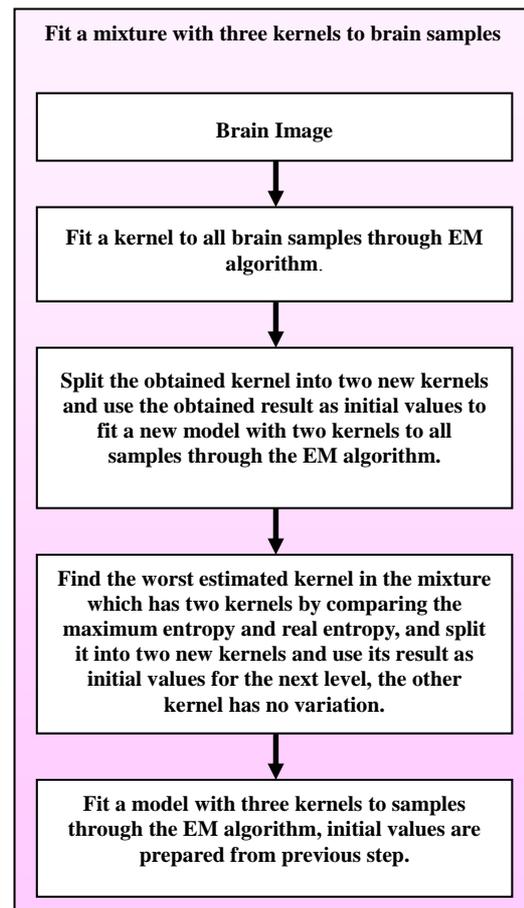


Figure 1. The Entropy-based EM algorithm to find the best fitted model to brain samples.

Assuming that the unobserved random field $\omega(x, y)$ is a Markov random field with the probability density function of $\omega(x, y)$ of a segmented image depending on its finite neighboring region $N(x, y)$ then

$$p(\omega(x, y) | \omega(x_i, y_i), \forall (x_i, y_i) \in N(x, y)) = p(\omega(x, y) | N(x, y)) \quad (27)$$

where $N(x, y)$ is a set of all labeled neighbors. So, the segmentation problem defined as a pixel classification problem using MAP can be estimated using a Gibbs distribution [22], which is easier to estimate. The Gibbs distribution can be expressed as

$$p(\omega(x, y) | N(x, y)) = \frac{1}{z} \exp^{-\frac{1}{T}U(\omega(x, y))} \quad (28)$$

where, $z = \sum_{\omega(x, y)} \exp^{-\frac{1}{T}U(\omega(x, y))}$ is a normalization constant which guarantees that $p(\omega(x, y))$ is always smaller or equal to one, and T is a constant which stands for the temperature constant which normally supposes to be one. And the energy function

$$U(\omega(x, y)) = \sum_{\omega(x_i, y_i) \in C(x, y)} V(\omega(x_i, y_i)) \quad (29)$$

where, cliques $C(x, y)$ are subsets of $N(x, y)$ or $\omega(x, y)$ itself and $V(\omega(x_i, y_i))$ is an arbitrary function of $\omega(x, y)$. The equivalence between MRF and Gibbs distribution is expressed by Hammersley-Clifford theorem, which states that $\omega(x, y)$ is a MRF with neighborhoods $N(x, y)$ if and only if $\omega(x, y)$ is a Gibbs distribution field with the cliques $C(x, y)$ induced by the neighborhood $N(x, y)$. This theorem provides an easy way to construct the MRF in an explicit manner, *i.e.*, one can explicitly estimate the conditional probability distribution of MRF by choosing specific kinds of cliques $C(x, y)$ and an appropriate energy function $U(\omega(x, y))$ that is specific for the practical problem. In our model we used the simple equation for the $U(\omega)$ energy function proposed by Therrien [23], and utilized by Nett *et al.* [24]. This equation is a linear combination of products of elements in the cliques:

$$U(\omega(x, y)) = \omega(x, y) (\alpha + \beta_1 (\omega(x-1, y) + \omega(x+1, y)) + \beta_2 (\omega(x, y-1) + \omega(x, y+1))) \quad (30)$$

9. COMPUTATION OF THE APRIORI PROBABILITY BY MRF MODEL

At first, the brain image is classified through the Bayesian classification method by using the model which is obtained as the result of Entropy Based EM algorithm. The gray level values for each tissue class resulted from

previous step are extracted. Then, the class conditional probability density function of each tissue class (*i.e.*, $\hat{\mu}_k(\omega)$ and $\hat{\sigma}_k^2(\omega)$) is estimated through the EM algorithm. Also, *a priori* probabilities ($P_k(\omega)$) for tissue classes are estimated by MRF model stated in Eqs.28-30. The parameters of MRF model, *i.e.*, α , β_1 and β_2 have been experimentally set to 0.1, 0.01, and 0.01 for the best result. Next, the new value for termination tolerance is calculated and evaluated. The new value of termination tolerance is calculated from

$$tol = \max \begin{pmatrix} \max(abs(\hat{P}_k(\omega) - \hat{P}_{k-1}(\omega))), \\ \max(abs(\hat{\mu}_k(\omega) - \hat{\mu}_{k-1}(\omega))), \\ \max(abs(\hat{\sigma}_k(\omega) - \hat{\sigma}_{k-1}(\omega))), \end{pmatrix} \quad (31)$$

in which $\hat{P}_k(\omega)$, $\hat{\mu}_k(\omega)$ and $\hat{\sigma}_k(\omega)$ are the row vectors of the *a priori* probabilities, mean, and variance of the tissue classes in k th iteration, respectively. This algorithm which computes the *a priori* probabilities by MRF model is shown in **Figure 2**.

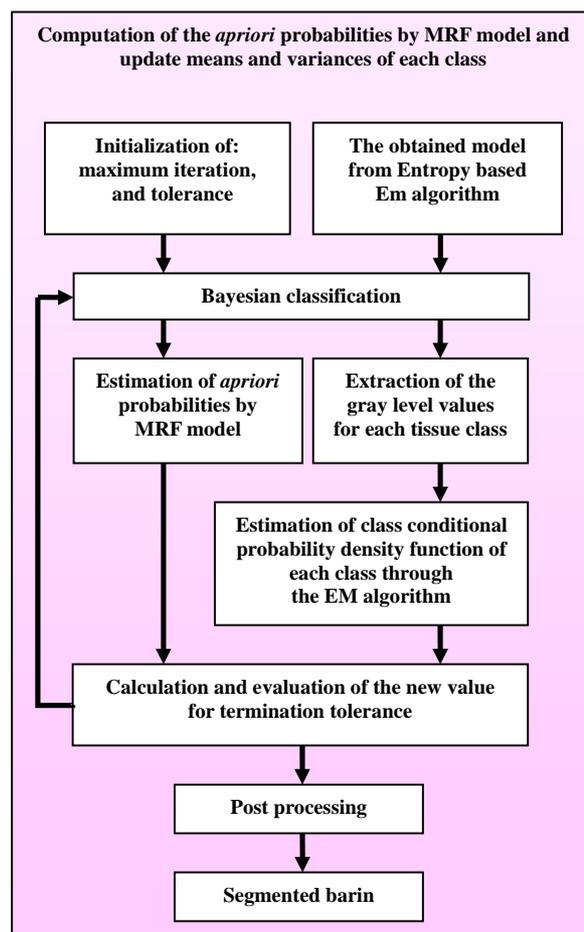


Figure 2. Computation of *a priori* probabilities and update of means and variances.

10. POSTPROCESSING

In manual segmentation, the possible lesions with sizes as small as one or two pixels are not usually considered as MS lesions by experts. However, as the gray level of neighboring pixels in an image usually are highly correlated, in our approach, the one-pixel object is considered as a possible noise if its gray level is higher than the average gray levels of its neighbor pixels (in 3×3 neighborhood) plus a margin of 25 (which is selected by trial). This object is recognized as noise and removed in the next step. Regarding the two pixel objects, in contrast with the manual procedure, these objects may be identified as seed points of MS lesions as recommended by the neurologist.

11. SIMULATION RESULTS

The results of proposed method, for two slices which

contain different lesion loads are shown in **Figure 3** and **Figure 4**. Segmented brain of typical original FLAIR images are shown in **Figure 3(a)** and **Figure 4(a)**. Multiple sclerosis pixels and the resulting segmented slices are shown in **Figure 3(b)** and **Figure 4(b)**. **Figure 3(c)** and **Figure 4(c)** show the histogram of brain pixels and estimated distribution through the proposed method. Also, three estimated kernels which constitute gaussian mixture model are shown in **Figure 3(d)** and **Figure 4(d)**.

12. EVALUATION

To evaluate the proposed method, SI [25], OF and EF [26] criteria are considered and computed over all 20 patients.

$$SI = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (32)$$

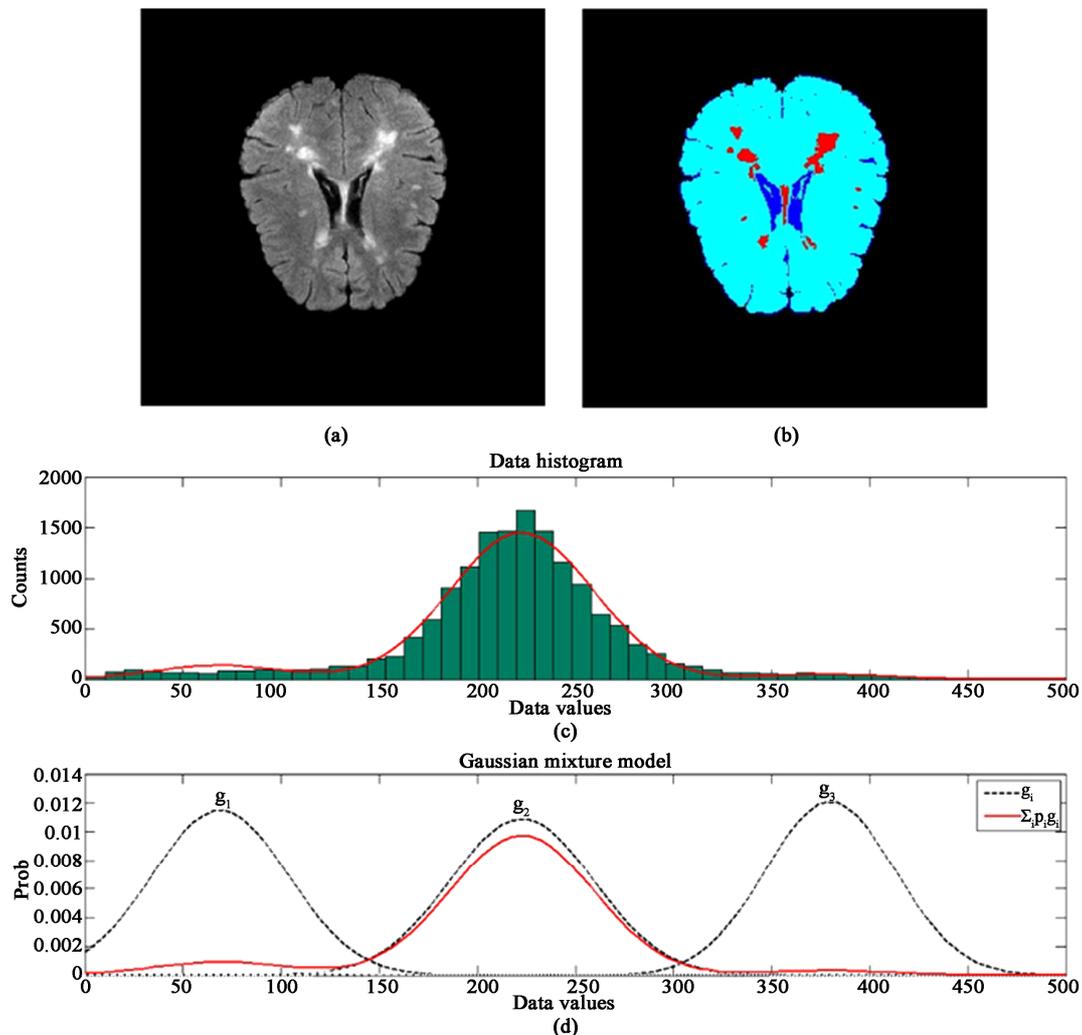


Figure 3. Result of applying the proposed algorithm to the image of a patient: (a) brain image; (b) result of proposed method; (c) histogram of brain, overlaid.

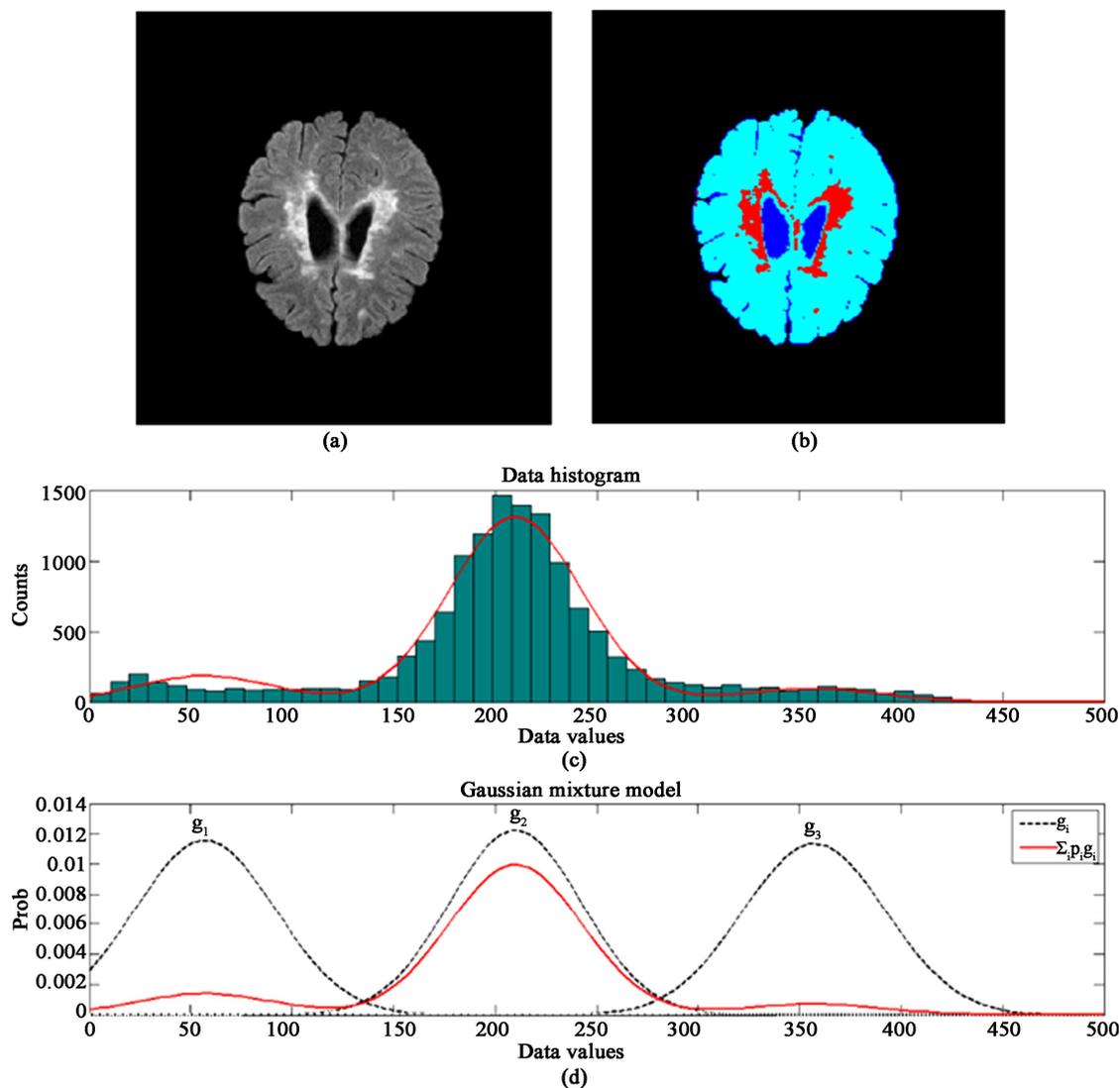


Figure 4. Result of applying the proposed algorithm to the image of a patient: (a) brain image; (b) result of proposed method; (c) histogram of brain, overlaid.

$$OF = \frac{TP}{TP + FN} \quad (33)$$

$$EF = \frac{FP}{TP + FN} \quad (34)$$

where, TP stands for true positive voxels, FP for false positive voxels, and FN for false negative voxels. SI and OF for a good segmentation should be close to 1 and EF should be close to 0. As mentioned before, patients are categorized into 3 groups. The proposed method has been applied to all patients and SI, OF and EF criteria are computed and mean values of each group are illustrated in **Table 1**. It is noticeable that SI, OF and EF are improved with an increase in lesion load. Also, the volumetric comparison of lesions between the proposed method and gold standard using correlation coefficient

(CC) for each patient group is shown in **Table 1**. By attention to correlation coefficient values, accuracy of the proposed method is increased for patients with large lesion load.

13. CONCLUSIONS

In this paper, we have described and validated a fully automatic method for classification of brain tissues in MR FLAIR images of MS patients. A gaussian mixture model which has three kernels as brain tissues is obtained through entropy based EM algorithm. This task starts with only one kernel and finally finds three kernels with no initially supplied training samples. Then, *a priori* probabilities of classes are estimated using MRF. At last, brain tissues are classified through Bayesian Classification. This framework has been validated on a data set of

Table 1. Similarity criteria and volumetric comparison of lesions for each patient group.

Patient category	N	Similarity criteria			Correlation analysis		
		SI	OF	EF	$M_{GS} \pm SD_{GS} (CC)$	$M \pm SD (CC)$	Correlation Coefficient
Small lesion load	7	0.7262	0.7277	0.2723	1.89 ± 1.13	1.87 ± 1.11	0.93
Moderate lesion load	10	0.7531	0.7395	0.2201	11.59 ± 3.10	11.58 ± 3.08	0.95
Large lesion load	3	0.8101	0.7857	0.1541	22.21 ± 3.54	22.23 ± 3.52	0.98

N: number of patients in each group, M: mean, SD: standard deviation

Table 2. Similarity index (SI) values for the proposed method and the other methods.

Method	Similarity index (SI)
Johnston <i>et al.</i> [3]	$\cong 0.65$
Boudraa <i>et al.</i> [13]	$\cong 0.62$
Leemput <i>et al.</i> [14]	$\cong 0.51$
Zijdenbos <i>et al.</i> [31]	$\cong 0.68$
Proposed Method	$\cong 0.75$

MR FLAIR images of 20 MS patients via similarity criteria (*i.e.*, SI, OF, and EF). There are other segmentation methods which have been evaluated via similar ways (*i.e.*, SI), and have used manual segmentation for evaluation of them, such as Johnston *et al.* [27], Boudraa *et al.* [28], Leemput *et al.* [29], and Zijdenbos *et al.* [30]. Comparison of proposed method with these methods is shown in **Table 2**. Johnston *et al.* [31] presented a semi-automatic segmentation approach based on stochastic relaxation method (SRM) and iterated conditional mode. Boudraa *et al.* [28] performed an automatic detection of MS lesions by using fuzzy C-means (FCM) algorithm. Leemput *et al.* [29] proposed the use of an intensity-based and stochastic model-aided segmentation approach for fully automatic segmentation of MS lesions. Zijdenbos *et al.* [30] developed an automatic pipeline based on a supervised artificial neural network (ANN) classifier. As it is seen in **Table 2**, the proposed method has improved the results reported in [26-29]. Also, in [31], it is mentioned that a value for SI more than 0.7 represents a very good segmentation in this field. Future work will include improving accuracy and correlation value.

REFERENCES

- [1] Kurtzke, J.F. (1983) Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, **33**, 1444-1452.
- [2] Rasoul, K., Mansur, V., Farzad, T., Massood Nabavi., S. (2008) Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. *Computers in Biology and Medicine*, **38**, 379-390.
- [3] Polman, C.H., Reingold, S.C., Edan, G., Fillippi, M., Hartung, H.P. and Kappos, L. (2005) Diagnostic criteria for MS 2005 revisions to the MC Donald criteria. *Annals of Neurology*, **58**, 840-846. [doi:10.1002/ana.20703](https://doi.org/10.1002/ana.20703)
- [4] Edelman, R.R., Hesselink, J.R. and Zlatkin, M.B. (2006) Clinical magnetic resonance imaging. 3rd Edition, Saunders, Philadelphia, 1571-1615.
- [5] Anbeek, P., Vincken, K.L., van Osch, M.J.P., Bisschops, R.H.C. and Van der Grond, J. (2004) Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage*, **21**, 1037-1044. [doi:10.1016/j.neuroimage.2003.10.012](https://doi.org/10.1016/j.neuroimage.2003.10.012)
- [6] Khayati, R. (2006) Quantification of multiple sclerosis lesions based on fractal analysis. PhD Thesis, Technical Report No. 1: Fully automatic object oriented brain segmentation in MRI. Amirkabir University of Technology, Tehran.
- [7] Duda, R.O., Hart, P.E. and Stork, D.G. (2001) Pattern classification. 2nd Edition, Wiley, New York.
- [8] Bernardo, J.M. and Smith, A.F.M. (1994) Bayesian theory. Wiley, New York. [doi:10.1002/9780470316870](https://doi.org/10.1002/9780470316870)
- [9] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1997) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1-38. [doi:10.1111.133.4884](https://doi.org/10.1111.133.4884)
- [10] Shannon, C. (1948) A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.
- [11] Cover, T.M. and Thomas, J.A. (1991) Elements of information theory. Wiley & Sons, New York. [doi:10.1002/0471200611](https://doi.org/10.1002/0471200611)
- [12] Rényi, A. (1961) On measures of entropy and information. University California Press, Berkeley, 547-561.
- [13] Leonenko, N., Pronzato, L. and Savani, V. (2008) A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, **36**, 2153-2182. [doi:10.1214/07-AOS539](https://doi.org/10.1214/07-AOS539)
- [14] Kozachenko, L. and Leonenko, N. (1987) On statistical estimation of entropy of a random vector. *Problems Information Transmission*, **23**, 95101.
- [15] Jones, M. and Sibson, R. (1987) What is projection pursuit. *Journal of the Royal Statistical Society: Series A*, **150**, 136. <http://www.jstor.org/stable/2981662>
- [16] Benavent, A.P. Ruiz, F.E. and Sáez, J.M. (2009) Learning Gaussian mixture models with entropy-based criteria. *IEEE Transactions on Neural Networks*, **20**, 1756-1771. [doi:10.1109/TNN.2009.2030190](https://doi.org/10.1109/TNN.2009.2030190)
- [17] Richardson, S. and Green, P. (1991) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society:*

- Series B*, **59**, 731-792. [doi:10.1111/1467-9868.00095](https://doi.org/10.1111/1467-9868.00095)
- [18] Dellaportas, P. and Papageorgiou, I. (2006) Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, **16**, 57-68. [doi:10.1007/s11222-006-5338-6](https://doi.org/10.1007/s11222-006-5338-6)
- [19] Mitchell, T.M. (1997) *Machine Learning*. McGraw-Hill, Boston. [doi:10.1036/0070428077](https://doi.org/10.1036/0070428077)
- [20] Li, S.Z. (2001) *Markov random field modeling in image analysis*. Springer, Tokyo.
- [21] Held, K., Kops, E.R., Krause, B.J., Wells, W.M., Kikinis, R. and Miller-Grtner, H.W. (1997) Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, **16**, 878-886. [doi:10.1109/42.650883](https://doi.org/10.1109/42.650883)
- [22] Geman, D., Geman, S., Graffigne, C. and Dong, P. (1990) Boundary detection by constrained optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-12**, 609-628. [doi:10.1109/34.56204](https://doi.org/10.1109/34.56204)
- [23] Therrien, C.W. (1989) *Decision, estimation, and classification*. Wiley, New York.
- [24] Nett, J.M. (2001) *The study of MS using MRI, image processing, and visualization*. Master's Thesis, University of Louisville, Louisville.
- [25] Zijdenbos, A.P., Dawant, B.M., Margolin, R.A. and Palmer, A.C. (1994) Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Transactions on Medical Imaging*, **13**, 716-724. [doi:10.1109/42.363096](https://doi.org/10.1109/42.363096)
- [26] Stokking, R., Vincken, K.L. and Viergever, M.A. (2000) Automatic morphologybased brain segmentation (MBRASE) from MRI-T1 data. *NeuroImage*, **12**, 726-738. [doi:10.1006/nimg.2000.0661](https://doi.org/10.1006/nimg.2000.0661)
- [27] Johnston, B., Atkins, M.S., Mackiewicz, B. and Anderson, M. (1996) Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, **15**, 154-169. [doi:10.1109/42.491417](https://doi.org/10.1109/42.491417)
- [28] Boudraa, A.O., Dehakb, S.M.R., Zhu, Y.M., Pachai, C., Bao, Y.G. and Grimaud, J. (2000) Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering. *Computers in Biology and Medicine*, **30**, 23-40. [doi:10.1016/S0010-4825\(99\)00019-0](https://doi.org/10.1016/S0010-4825(99)00019-0)
- [29] Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A. and Suetens, P. (2001) Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, **20**, 677-688. [doi:10.1109/42.938237](https://doi.org/10.1109/42.938237)
- [30] Zijdenbos, A.P., Forghani, R. and Evans, A.C. (2002) Automatic pipeline analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE Transactions on Medical Imaging*, **21**, 1280-1291. [doi:10.1109/TMI.2002.806283](https://doi.org/10.1109/TMI.2002.806283)
- [31] Bartko, J.J. (1991) Measurement and reliability: statistical thinking considerations. *Schizophrenia Bulletin*, **17**, 483-489. [doi:10.1093/schbul/17.3.483](https://doi.org/10.1093/schbul/17.3.483)